
Andrea Montanari

with David Donoho, Arian Maleki, Mohsen Bayati and Jose Bento

May 27, 2010
What is this talk about?

\[ y = A x_0 + w \]

Estimate \( x_0 \in \mathbb{R}^N \) given \((y, A)\).
What is this talk about?

\[ y = A x_0 + w \]

Estimate \( x_0 \in \mathbb{R}^N \) given \((y, A)\).
What is this talk about?

$$y = A x_0 + w$$

Estimate $x_0 \in \mathbb{R}^N$ given $(y, A)$. 
Problem: $x_0$ is simple/structured (e.g. sparse)

$\rightarrow$ Good estimators: optimal MSE + low complexity.

A general statistical theory is possible.
Problem: $x_0$ is simple/structured (e.g. sparse)

$\rightarrow$ Good estimators: optimal MSE + low complexity.

A general statistical theory is possible.
1. Define a convex cost $C_{A,y}(x)$.

2. Estimate

$$\hat{x}(y, A) = \arg\min_{x \in \mathbb{R}^N} C_{A,y}(x)$$
The ‘Stanford recipe’

1. Define a convex cost $C_{A,y}(x)$.

2. Estimate

$$\hat{x}(y, A) = \arg\min_{x \in \mathbb{R}^N} C_{A,y}(x)$$
1. Define a convex cost $C_{A,y}(x)$.

2. Estimate

$$\hat{x}(y, A) = \arg\min_{x \in \mathbb{R}^N} C_{A,y}(x)$$
The LASSO

\[ \hat{x}(y, A) = \arg\min_{x \in \mathbb{R}^N} C_{A,y}(x) \]

\[ C_{A,y}(x) = \lambda \|x\|_1 + \frac{1}{2} \|y - Ax\|_2^2 \]

[Tibshirani 96; Chen, Donoho 95]
Wonderful, but...

→ What performance should I expect?

→ How am I supposed to choose $C_{A,y}$?

→ What if I can design $A$?
Amuse-bouche

\[ A \rightarrow \text{‘real’ data} \]

\[ x_{0,i} = \begin{cases} 
+1 & \text{with prob. 0.064}, \\
0 & \text{with prob. 0.872}, \\
-1 & \text{with prob. 0.064}, 
\end{cases} \]

\[ w_i \sim \mathcal{N}(0, 0.2) \]
Clinical data

\[ A \text{ is } n \times N, \ n = 0.64N \]
Clinical data

$A$ is $n \times N$, $n = 0.64N$
Gene expression data

$A$ is $85 \times 200$ [from Hastie, Tibshirani, Friedman]
A is $85 \times 200$ [from Hastie, Tibshirani, Friedman]
Outline

1. Linear estimation
2. Non-linear estimation: The scalar case
3. AMP: Iterative non-linear estimation
4. Why is this exciting ?!
5. Proof sketch (?)

arXiv:0907.3574
arXiv:1001.3448
arXiv:1004.1218
Setting

$$y = Ax_0 + w,$$

$$x_0 \in \mathbb{R}^N, \; y, w \in \mathbb{R}^n, \; A \in \mathbb{R}^{n \times N}$$
Setting

\[ y = Ax_0 + w, \]

\[ x_0 \in \mathbb{R}^N, \quad y, w \in \mathbb{R}^n, \quad A \in \mathbb{R}^{n \times N} \]
Setting

\[ y = A x_0 + \mathcal{N} \]
Setting: Normalization

\[
\rightarrow w \sim \mathcal{N}(0, \sigma^2 I_{n \times n})
\]

\[
\rightarrow N, n \to \infty, \ n/N = \delta
\]

\[
\rightarrow A = [A_1 | \cdots | A_N] \quad ||A_i||_2 \approx 1
\]
Linear estimation
Linear estimation

\[ y = Ax_0 + w, \quad \mathbb{E}[x_0 x_0^T] \preceq I. \]

The estimator

\[ \eta_{\text{lin}}(y) = (A^T A + \sigma^2 I)^{-1} A^T y. \]

\( \eta_{\text{lin}} \) is minimax optimal (over the class \( \{ x_0 | \mathbb{E}[x_0 x_0^T] \preceq I \} \))
Linear estimation

\[ y = Ax_0 + w, \quad \mathbb{E}[x_0 x_0^T] \preceq I. \]

The estimator

\[ \eta_{\text{lin}}(y) = (A^T A + \sigma^2 I)^{-1} A^T y. \]

\( \eta_{\text{lin}} \) is minimax optimal (over the class \{\( x_0 | \mathbb{E}[x_0 x_0^T] \preceq I \})
Linear estimation

\[ y = Ax_0 + w, \quad \mathbb{E}[x_0 x_0^T] \preceq I. \]

The estimator

\[ \eta_{\text{lin}}(y) = (A^T A + \sigma^2 I)^{-1} A^T y. \]

\( \eta_{\text{lin}} \) is minimax optimal (over the class \( \{ x_0 | \mathbb{E}[x_0 x_0^T] \preceq I \} \)).
A naive iterative scheme

\[ \eta_{\text{lin}}(y) = (A^T A + \sigma^2 I)^{-1} A^T y. \]

\[ \begin{align*}
\sigma^2 x + A^T A x &= A^T x_0, \\
(1 + \sigma^2) x &= A^T y - (A^T A - I) x, \\
x^{t+1} &= \frac{1}{1 + \sigma^2} \{ A^T y - (A^T A - I) x^t \},
\end{align*} \]

Something looks weird…
A naive iterative scheme

\[ \eta_{\text{lin}}(y) = (A^T A + \sigma^2 I)^{-1} A^T y . \]

\[ \sigma^2 x + A^T A x = A^T x_0 , \]
\[ (1 + \sigma^2) x = A^T y - (A^T A - I) x , \]
\[ x^{t+1} = \frac{1}{1 + \sigma^2} \{ A^T y - (A^T A - I) x^t \} , \]

\[ x^{t+1} = \frac{1}{1 + \sigma^2} (x^t + A^T z^t) , \]
\[ z^t = y - A x^t . \]

Something looks weird…
A naive iterative scheme

$$\eta_{\text{lin}}(y) = (A^T A + \sigma^2 I)^{-1} A^T y .$$

$$\sigma^2 x + A^T A x = A^T x_0 ,$$

$$(1 + \sigma^2) x = A^T y - (A^T A - I) x ,$$

$$x^{t+1} = \frac{1}{1 + \sigma^2} \left\{ A^T y - (A^T A - I) x^t \right\} ,$$

$$x^{t+1} = \frac{1}{1 + \sigma^2} (x^t + A^T z^t) ,$$

$$z^t = y - Ax^t .$$

Something looks weird…
A naive iterative scheme

\[\eta_{\text{lin}}(y) = (A^T A + \sigma^2 I)^{-1} A^T y.\]

\[
\begin{align*}
\sigma^2 x + A^T Ax &= A^T x_0, \\
(1 + \sigma^2) x &= A^T y - (A^T A - I) x, \\
x^{t+1} &= \frac{1}{1 + \sigma^2} \{A^T y - (A^T A - I)x^t\}, \\
x^{t+1} &= \frac{1}{1 + \sigma^2} (x^t + A^T z^t), \\
z^t &= y - Ax^t.
\end{align*}
\]

Something looks weird...
A naive iterative scheme

\[ \eta_{\text{lin}}(y) = (A^T A + \sigma^2 I)^{-1} A^T y. \]

\[
\begin{align*}
\sigma^2 x + A^T A x & = A^T x_0, \\
(1 + \sigma^2) x & = A^T y - (A^T A - I) x, \\
x^{t+1} & = \frac{1}{1 + \sigma^2} \{ A^T y - (A^T A - I) x^t \},
\end{align*}
\]

Something looks weird...
Non-linear estimation: The scalar case
Sparse signal ($N = n = 1$)

\[ y = x_0 + w, \quad \mathbb{P}\{x_0 \neq 0\} \leq \epsilon. \]

The estimator

\[ \eta(y; \theta) = \begin{cases} 
  y - \theta & \text{for } \theta < y, \\
  0 & \text{for } -\theta \leq y \leq \theta, \\
  y + \theta & \text{for } y < -\theta.
\end{cases} \]

\eta is minimax optimal, for \(\theta = \tau(\epsilon) \sigma\)
(over the class \(\{x_0| \mathbb{P}\{x_0 \neq 0\} \leq \epsilon\}\))

[Donoho, Johnstone, 1994]
Sparse signal \((N = n = 1)\)

\[ y = x_0 + w, \quad \mathbb{P}\{x_0 \neq 0\} \leq \epsilon. \]

The estimator

\[ \eta(y; \theta) = \begin{cases} 
    y - \theta & \text{for } \theta < y, \\
    0 & \text{for } -\theta \leq y \leq \theta, \\
    y + \theta & \text{for } y < -\theta. 
\end{cases} \]

\eta is minimax optimal, for \(\theta = \tau(\epsilon) \sigma\) (over the class \(\{x_0 | \mathbb{P}\{x_0 \neq 0\} \leq \epsilon\}\))

[Donoho, Johnstone, 1994]
Sparse signal \((N = n = 1)\)

\[ y = x_0 + w, \quad \mathbb{P}\{x_0 \neq 0\} \leq \epsilon. \]

The estimator

\[ \eta(y; \theta) = \begin{cases} 
    y - \theta & \text{for } \theta < y, \\
    0 & \text{for } -\theta \leq y \leq \theta, \\
    y + \theta & \text{for } y < -\theta. 
\end{cases} \]

\(\eta\) is minimax optimal, for \(\theta = \tau(\epsilon) \sigma\)

(over the class \(\{x_0 | \mathbb{P}\{x_0 \neq 0\} \leq \epsilon\}\))

[Donoho, Johnstone, 1994]
\[ \eta(y; \theta) \]
AMP: Iterative non-linear estimation
Non-linear estimation

\[ y = Ax_0 + w, \quad \mathbb{P}\{x_0,i \neq 0\} \leq \epsilon. \]

Cannot use linear algebra but . . .

\[
\begin{align*}
    x^{t+1} &= \frac{1}{1 + \sigma^2} (x^t + A^T z^t), \\
    z^t &= y - Ax^t.
\end{align*}
\]
Non-linear estimation

\[ y = Ax_0 + w, \quad \mathbb{P}\{x_0,i \neq 0\} \leq \epsilon. \]

Cannot use linear algebra but . . .

\[
\begin{align*}
    x^{t+1} &= \frac{1}{1 + \sigma^2}(x^t + A^Tz^t), \\
    z^t &= y - Ax^t.
\end{align*}
\]
Non-linear estimation

\[ y = Ax_0 + w, \quad \mathbb{P}\{x_0, i \neq 0\} \leq \epsilon. \]

Cannot use linear algebra but...

\[
\begin{align*}
    x^{t+1} &= \eta(x^t + A^T z^t; \theta_t), \\
    z^t &= y - Ax^t.
\end{align*}
\]

[IST: Daubechies, Defrise, De Mol, 2004]
There is no reason for this to work :-(

\[ x^{t+1} = \eta(x^t + A^T z^t; \theta_t) , \]
\[ z^t = y - Ax^t . \]

There is substantial evidence that this works* :-) 

\[ x^{t+1} = \eta(x^t + A^T z^t; \theta_t) , \]
\[ z^t = y - Ax^t + b_t z^{t-1} . \] (AMP)

with
\[ b_t \equiv \frac{1}{n} \sum_{i=1}^{N} \eta'(x^{t-1} + A^T z^{t-1}; \theta_t) . \]
There is no reason for this to work  

$$x^{t+1} = \eta(x^t + A^T z^t; \theta_t),$$

$$z^t = y - Ax^t.$$  

There is substantial evidence that this works* :-)

$$x^{t+1} = \eta(x^t + A^T z^t; \theta_t),$$

$$z^t = y - Ax^t + b_t z^{t-1}.$$  \hspace{1cm} (AMP)

with  

$$b_t \equiv \frac{1}{n} \sum_{i=1}^{N} \eta'(x^{t-1} + A^T z^{t-1}; \theta_t).$$
There is no reason for this to work  :-(

\[ x^{t+1} = \eta(x^t + A^T z^t; \theta_t) , \]
\[ z^t = y - A x^t. \]

There is substantial evidence that this works*  :-)  

\[ x^{t+1} = \eta(x^t + A^T z^t; \theta_t) , \quad (\text{AMP}) \]
\[ z^t = y - A x^t + b_t z^{t-1} . \]

with  
\[ b_t \equiv \frac{1}{n} \sum_{i=1}^{N} \eta'(x^{t-1} + A^T z^{t-1}; \theta_t) . \]
What's the big deal with $+ b_t z^{t-1}$ ('Onsager term')?

Distribution of $(x^t + A^T z^t)_i$ conditional on $x_0,i = 1$
A theorem

**Theorem (Bayati, Montanari, 2010)**

Assume $A_{ij} \sim \mathcal{N}(0, 1/n)$; empirical law of $x_0$ converges. Then,

$$
\text{empirical law of } \{(x^t + A^T z^t)_i - x_{0,i}\} \xrightarrow{w} \mathcal{N}(0, \tau_t)
$$

almost surely as $n \to \infty$.

with

$$
\tau_{t+1}^2 = \sigma^2 + \frac{1}{\delta} \mathbb{E}\{[\eta(X_0 + \tau_t Z; \theta_t) - X_0]^2\}
$$

[State Evolution]
A theorem

Theorem (Bayati, Montanari, 2010)

Assume $A_{ij} \sim N(0, 1/n)$; empirical law of $x_0$ converges. Then,

$$\text{empirical law of } \{(x^t + A^T z^t)_i - x_0,i\} \overset{w}{\Rightarrow} N(0, \tau_t)$$

almost surely as $n \to \infty$.

with

$$\tau_{t+1}^2 = \sigma^2 + \frac{1}{\delta} \mathbb{E}\{[\eta(X_0 + \tau_t Z; \theta_t) - X_0]^2\}$$

[State Evolution]
A theorem

**Theorem (Bayati, Montanari, 2010)**

Assume $A_{ij} \sim \mathcal{N}(0, 1/n)$; empirical law of $x_0$ converges. Then,

$$\text{empirical law of } \{(x^t + A^T z^t)_i - x_0, i\} \xrightarrow{w} \mathcal{N}(0, \tau_t)$$

almost surely as $n \rightarrow \infty$.

with

$$\tau_{t+1}^2 = \sigma^2 + \frac{1}{\delta} \mathbb{E}\{[\eta(X_0 + \tau_t Z; \theta_t) - X_0]^2\}$$

**State evolution $\approx$ Density evolution for dense graphs**

Andrea Montanari (Stanford)
Universality: Random Fourier Ensemble

\[ A^T = [R_1 | \ldots | R_n], \quad (R_i x_0) = \text{Random Fourier coefficient}. \]
Another theorem

Theorem (Bayati, Montanari, 2010)

Assume $A_{ij} \sim \mathcal{N}(0, 1/n)$, $y = Ax_0 + w$, and $(\tau_\infty^2, \theta_\infty) \text{ unique solution of}$

$$
\tau_\infty^2 = \sigma^2 + \frac{1}{\delta} \mathbb{E}\{[\eta(X_0 + \tau_\infty Z; \theta_\infty) - X_0]^2\},
$$

$$
\lambda = \theta_\infty \{1 - \frac{1}{\delta} \mathbb{E}[\eta'(X_0 + \tau_\infty Z; \theta_\infty)]\}
$$

Then,

$$
\lim_{N \to \infty} \frac{1}{N} \|\hat{x}_{\text{LASSO}}(\lambda) - x_0\|^2 = (\tau_\infty^2 - \sigma^2)\delta.
$$

almost surely as $n \to \infty$. 

Conjectured in a more general context with Donoho and Maleki
Another theorem

**Theorem (Bayati, Montanari, 2010)**

Assume $A_{ij} \sim \mathcal{N}(0, 1/n)$, $y = Ax_0 + w$, and $(\tau_\infty^2, \theta_\infty)$ unique solution of

\[
\tau_\infty^2 = \sigma^2 + \frac{1}{\delta} \mathbb{E}\{[\eta(X_0 + \tau_\infty Z; \theta_\infty) - X_0]^2\},
\]

\[
\lambda = \theta_\infty \{1 - \frac{1}{\delta} \mathbb{E}[\eta'(X_0 + \tau_\infty Z; \theta_\infty)]\}
\]

Then,

\[
\lim_{N \to \infty} \frac{1}{N} \| \hat{x}_{\text{LASSO}}(\lambda) - x_0 \|^2 = (\tau_\infty^2 - \sigma^2)\delta.
\]

*almost surely as $n \to \infty$.*

Conjectured in a more general context with Donoho and Maleki
Why is this exciting ?!!
Connection with compressed sensing?

$$y = Ax_0 + w$$

Andrea Montanari (Stanford)
Three phase diagrams \((n, N \to \infty)\)

\[ \frac{n}{N} \to \delta \]

\[ \frac{1}{N} \| x_0 \|_0 \to \rho \delta \]
Three phase diagrams \((n, N \to \infty)\)

\[ \frac{n}{N} \to \delta \]

\[ \frac{1}{N} \| x_0 \|_0 \to \rho \delta \]
The ‘golden standard’

Basis pursuit (noiseless)

\[
\begin{align*}
\text{minimize} & \quad \|x\|_1, \\
\text{subject to} & \quad y = Ax.
\end{align*}
\]

LASSO (noisy)

\[
\begin{align*}
\text{minimize} & \quad \lambda \|x\|_1 + \frac{1}{2} \|y - Ax\|^2.
\end{align*}
\]
The ‘golden standard’

Basis pursuit (noiseless)

\[
\begin{align*}
& \text{minimize} & \|x\|_1, \\
& \text{subject to} & y = Ax.
\end{align*}
\]

LASSO (noisy)

\[
\begin{align*}
& \text{minimize} & \lambda \|x\|_1 + \frac{1}{2} \|y - Ax\|^2.
\end{align*}
\]
The ‘golden standard’

Basis pursuit (noiseless)

\[
\begin{align*}
\text{minimize} & \quad \|x\|_1, \\
\text{subject to} & \quad y = Ax.
\end{align*}
\]

LASSO (noisy)

\[
\begin{align*}
\text{minimize} & \quad \lambda \|x\|_1 + \frac{1}{2} \|y - Ax\|^2.
\end{align*}
\]
Phase diagram 1: \( \ell_0-\ell_1 \) equivalence (noiseless)

\[
N, n \to \infty, \, n/N = \delta, \quad ||x_0||_0/N = \delta \rho
\]

\( \ell_1 \) reconstructs \( x_0 \) with high probability

\( \ell_1 \) fails

[Donoho, Tanner, 2006, based on Vershik, Sposhyev 1992]

Can be recovered from this techniques
Phase diagram 1: ‘\(l_0-l_1\) equivalence’ (noiseless)

\[ N, n \to \infty, \frac{n}{N} = \delta, \quad \|x_0\|_0/N = \delta \rho \]

\(l_1\) reconstructs \(x_0\) with high probability

\(l_1\) fails

[Donoho, Tanner, 2006, based on Vershik, Sposhyev 1992]

Can be recovered from this techniques
Phase diagram 2: Algorithm comparison (noiseless)

Comparison of Different Algorithms

IHT
IST
Tuned TST
LARS
OMP
L1
MPIST

δ
ρ

Andrea Montanari (Stanford)
Universality: $A$

Phase transition of FOAMP for different matrix ensembles

Theoretical $L_1$ AMP, USE AMP, Fourier AMP, Rademacher

Andrea Montanari (Stanford)
Phase diagram 3: LASSO minimax risk (noisy)

$$\inf_\lambda \sup_{X_0} \text{MSE}(X_0; \sigma^2) = M^* (\delta, \rho) \sigma^2$$
Proof sketch
Theorem I: A slightly simpler setting

\[ A \in \mathbb{R}^{n \times n} \]

\[
A_{ij} = \begin{cases} 
\mathcal{N}(0, 1/n) & \text{independent, if } i < j, \\
0 & \text{if } i = j, \\
A_{ji} & \text{if } i > j.
\end{cases}
\]

\[ x^{t+1} = Af(x^t) - b_t f(x^{t-1}) \]

\[ x^t \in \mathbb{R}^n, \quad x^0 = 0, \quad f : \mathbb{R} \to \mathbb{R} \]
Theorem I: A slightly simpler setting

\[ A \in \mathbb{R}^{n \times n} \]

\[ A_{ij} = \begin{cases} 
N(0, 1/n) & \text{independent, if } i < j, \\
0 & \text{if } i = j, \\
A_{ji} & \text{if } i > j.
\end{cases} \]

\[ x^t \in \mathbb{R}^n, \quad x^0 = 0, \quad f : \mathbb{R} \to \mathbb{R} \]

\[ x^{t+1} = Af(x^t) - b_t f(x^{t-1}) \]
Conditioning technique (Erwin Bolthausen)

\[ x^{t+1} = Af(x^t) - b_t f(x^{t-1}) \equiv \lambda^t - b t m^{t-1} \]

\[ \mathcal{G}_t \equiv \sigma(\{x^0, x^1, \ldots, x^t\}) \]

**Problem:** \( A \) and \( \mathcal{G}_t \) are dependent

**Idea:** Compute the distribution of \( A \) conditional on \( \mathcal{G}_t \).
Conditioning technique

\[ \mathcal{E}_t \equiv \{ x^1 + \lambda^0 m^{-1} = A m^0, \ldots, x^t + \lambda^{t-1} m^{t-2} = A m^{t-1} \} \]

\[ A|\mathcal{G}_t \overset{d}{=} \mathbb{E}\{A|\mathcal{G}_t\} + P^t A^{\text{new}} P^t \perp \]

\[ x^{t+1} = A m^t - b_t m^{t-1} \]

\[ \approx A^{\text{new}} m^t + E_t m^t - b_t m^{t-1} \]

\[ -b_t m^{t-1} \] cancels non-gaussian terms in \( E_t m^t \)!
Conditioning technique

\[ \mathcal{E}_t \equiv \{ x^1 + \lambda^0 m^{-1} = A m^0, \ldots, x^t + \lambda^{t-1} m^{t-2} = A m^{t-1} \} \]

\[
A|\mathcal{G}_t = A|\mathcal{E}_t \overset{d}{=} \mathbb{E}\{A|\mathcal{G}_t\} + P^t_\perp A^{\text{new}} P^t_\perp \\
\equiv E_t + P^t_\perp A^{\text{new}} P^t_\perp
\]

\[
x^{t+1} = A m^t - b_t m^{t-1} \\
= P^t_\perp A^{\text{new}} P^t_\perp m^t + E_t m^t - b_t m^{t-1} \\
\approx A^{\text{new}} m^t + E_t m^t - b_t m^{t-1}
\]

\(-b_t m^{-t-1}\) cancels non-gaussian terms in \(E_t m^t\)!
Conditioning technique

\[ \mathcal{E}_t \equiv \{ x_1 + \lambda^0 m^{-1} = Am^0, \ldots, x^t + \lambda^{t-1} m^{t-2} = Am^{t-1} \} \]

\[
A|\mathcal{S}_t \overset{d}{=} A|\mathcal{E}_t \equiv \mathbb{E}\{A|\mathcal{S}_t\} + P_t A^{\text{new}} P_t^\perp \\
\equiv \mathcal{E}_t + P_t A^{\text{new}} P_t^\perp
\]

\[ x^{t+1} = Am^t - b_t m^{t-1} \]

\[ \approx A^{\text{new}} m^t + E_t m^t - b_t m^{t-1} \]

\[-b_t m^{-t-1} \text{ cancels non-gaussian terms in } E_t m^t! \]
Conditioning technique

\[ \mathcal{E}_t \equiv \{ x^1 + \lambda^0 m^{-1} = Am^0, \ldots, x^t + \lambda^{t-1} m^{t-2} = Am^{t-1} \} \]

\[
A|_{\mathcal{G}_t} = A|_{\mathcal{E}_t} \overset{d}{=} \mathbb{E}\{A|_{\mathcal{G}_t}\} + P_\perp A^{\text{new}} P_\perp \\
\equiv E_t + P_\perp A^{\text{new}} P_\perp
\]

\[ x^{t+1} = Am^t - b_t m^{t-1} \]
\[ = P_\perp A^{\text{new}} P_\perp m^t + E_t m^t - b_t m^{t-1} \]
\[ \approx A^{\text{new}} m_\perp + E_t m^t - b_t m^{t-1} \]

\(-b_t m^{-t-1}\) cancels non-gaussian terms in \(E_t m^t\)!
Conditioning technique

\[ \mathcal{E}_t \equiv \{ x^1 + \lambda^0 m^{-1} = Am^0, \ldots, x^t + \lambda^{t-1} m^{t-2} = Am^{t-1} \} \]

\[
A|\mathcal{G}_t = A|\mathcal{E}_t \overset{d}{=} \mathbb{E}\{ A|\mathcal{G}_t \} + P^t A^{\text{new}} P^t \\
\equiv E_t + P^t A^{\text{new}} P^t
\]

\[
x^{t+1} = Am^t - b_t m^{t-1} \\
= P^t A^{\text{new}} P^t m^t + E_t m^t - b_t m^{t-1} \\
\approx A^{\text{new}} m^t + E_t m^t - b_t m^{t-1}
\]

\[-b_t m^{t-1} \text{ cancels non-gaussian terms in } E_t m^t! \]
Theorem II: Basic idea

\[ x^{t+1} = \eta(x^t + A^T z^t; \theta_t), \]
\[ z^t = y - Ax^t + b_t z^{t-1}. \]  

\textbf{Idea:} \hspace{1cm} x^t \rightarrow \arg\min_x C_{A,y}(x)
Implementation

1. Construct

\[ \text{sg}(x^t) \in \partial C_{A,y}(x^t). \]

2. Use state evolution to prove that

\[ \| \text{sg}(x^t) \|^2_2 \leq N e^{-c_1 t}. \]

3. Conclude that

\[ C_{A,y}(x^t) \leq \min_x C_{A,y}(x) + N e^{-c_2 t}. \]

We are not quite done, but...
**Implementation**

1. Construct

   \[ \text{sg}(x^t) \in \partial C_{A,y}(x^t). \]

2. Use state evolution to prove that

   \[ \|\text{sg}(x^t)\|_2^2 \leq N e^{-c_1 t}. \]

3. Conclude that

   \[ C_{A,y}(x^t) \leq \min_x C_{A,y}(x) + N e^{-c_2 t}. \]

We are not quite done, but...
Implementation

1. Construct

\[ \text{sg}(x^t) \in \partial C_{A,y}(x^t). \]

2. Use state evolution to prove that

\[ \| \text{sg}(x^t) \|_2^2 \leq N e^{-c_1 t}. \]

3. Conclude that

\[ C_{A,y}(x^t) \leq \min_x C_{A,y}(x) + N e^{-c_2 t}. \]

We are not quite done, but...
Implementation

1. Construct

\[ \text{sg}(x^t) \in \partial C_{A,y}(x^t). \]

2. Use state evolution to prove that

\[ \|\text{sg}(x^t)\|_2^2 \leq N e^{-c_1 t}. \]

3. Conclude that

\[ C_{A,y}(x^t) \leq \min_x C_{A,y}(x) + N e^{-c_2 t}. \]

We are not quite done, but...
Implementation

1. Construct

\[ \text{sg}(x^t) \in \partial C_{A,y}(x^t). \]

2. Use state evolution to prove that

\[ \|\text{sg}(x^t)\|_2^2 \leq N e^{-c_1 t}. \]

3. Conclude that

\[ C_{A,y}(x^t) \leq \min_x C_{A,y}(x) + N e^{-c_2 t}. \]

We are not quite done, but...
Conclusion
I did not talk about:

- Where does the Onsager term come from?
- Relation with convex optimization.
- Replicas: non-rigorous, non-algorithmic [Rangan, Fletcher, Goyal - Kabashima, Tanaka - Baron, Guo Shamai, 2009] ⇒ SE: rigorous, algorithmic

Thanks!
I did not talk about:

- Where does the Onsager term come from?
- Relation with convex optimization.
- Replicas: non-rigorous, non-algorithmic [Rangan, Fletcher, Goyal - Kabashima, Tanaka - Baron, Guo Shamai, 2009] ⇒ SE: rigorous, algorithmic

Thanks!
I did not talk about:

- Where does the Onsager term come from?
- Relation with convex optimization.

Replicas: non-rigorous, non-algorithmic [Rangan, Fletcher, Goyal - Kabashima, Tanaka - Baron, Guo Shamai, 2009]
⇒ SE: rigorous, algorithmic

Thanks!
I did not talk about:

- Where does the Onsager term come from?
- Relation with convex optimization.
- Replicas: non-rigorous, non-algorithmic [Rangan, Fletcher, Goyal - Kabashima, Tanaka - Baron, Guo Shamai, 2009]

⇒ SE: rigorous, algorithmic

Thanks!
I did not talk about:

- Where does the Onsager term come from?
- Relation with convex optimization.
- Replicas: non-rigorous, non-algorithmic [Rangan, Fletcher, Goyal - Kabashima, Tanaka - Baron, Guo Shamai, 2009]
  ⇒ SE: rigorous, algorithmic

Thanks!
More simulations
Universality: $s_0$

\[ \delta = 0.1 \]

\[ \delta = 0.3 \]
Comparison of Different Algorithms

- IHT
- IST
- Tuned TST
- LARS
- OMP
- L1
- MPIST

\[ \rho \] vs. \[ \delta \]

Andrea Montanari (Stanford)
Gaussian $A$

Andrea Montanari (Stanford)
$A_{ij} \in \{+1/\sqrt{n}, -1/\sqrt{n}\}$
A graph showing the relationship between $\text{MSE}$ and $\lambda$ for different values of $N$: $N=200$, $N=500$, and $N=1000$. The graph indicates that as $\lambda$ increases, the MSE decreases for all values of $N$. The $x$-axis represents $\lambda$ ranging from 0 to 2, and the $y$-axis represents the MSE ranging from 0 to 0.5.